

# Compte-rendu du Comité stratégique de RDA

Réunion du 13 septembre 2012

## **Étaient présents :**

Yves Alix (IGB)  
Philippe Beauvillard (Electre)  
Raymond Bérard (ABES, président du Comité stratégique)  
Emmanuelle Bermès (Centre Pompidou)  
Christophe Catanese (ENSSIB)  
Pierre Fuzeau (CG 46)  
Gildas Illien (BnF),  
Cécile Jallet (ADBBDP)  
Julie Ladant (MISTRD)  
Patricia Le Galèze (SLL, secrétaire de la réunion)  
Philippe Le Pape (ABES, rapporteur du groupe technique)  
Françoise Leresche (BnF, rapporteuse du groupe technique)  
Marc Maisonneuve (Tosca Consultants)  
Sophie Mazens (AURA)  
Grégory Miura (ADBU)  
Jenny Rigaud (CNFPT)  
Robert Zachayus (FULBI)

## **Excusés :**

Françoise Bourdon (BnF, rapporteuse du groupe technique)  
Fabien Plazannet (SLL)  
Dominique Lahary (IABD)  
ADBGV

## **Absents :**

ABF  
Sabine Donnard-Cusse (AFNOR)

## **Ordre du jour :**

- Présentation du rapport W3C « Library Linked Data » (Emmanuelle Bermès)
- Présentation du rapport du rapport d'orientation demandé par le Comité stratégique bibliographique

Prochaine réunion du Comité stratégique RDA : 18 janvier 2013, 14h-17h au Cercle de la Librairie

Ordre du jour :            Décisions du Comité stratégique bibliographique  
                                      Travaux du groupe technique

Raymond Bérard présente les nouveaux membres du Comité stratégique RDA :  
Yves Alix, Inspection générale des bibliothèques  
Julie Ladant, MISTRD, qui succède à Marie-Odile Illiano  
Jenny Rigaud, CNFPT Pôle culture – INSET de Nancy

## **1 Présentation du rapport W3C « Library Linked Data » (Emmanuelle Bermès)**

L'élaboration de ce rapport s'est faite de mai 2010 à août 2011 sous l'égide du W3C par un groupe d'incubation sur les bibliothèques et le web de données. Ce rapport a été publié en novembre 2011.

Avant de présenter le rapport, Emmanuelle Bermès présente les notions de Linked Data, Web de données, Web sémantique.

L'utilité du web de données est de permettre l'intégration des données structurées. Cette intégration de données structurées dans le web se fait par l'intermédiaire d'URI, des sortes d'identifiants valables sur le web, utilisables par un navigateur associé au protocole HTTP. L'association de ces deux technologies permet de s'inscrire dans le web.

La nouveauté pour les bibliothèques est l'utilisation des normes du web sémantique dédiées aux données structurées et se caractérisant par la création de liens. Le web sémantique correspond à l'ensemble de normes technologiques, le web de données à l'application des technologies aux données. Le web de données encourage donc l'utilisation des normes du W3C notamment RDF, le format du web de données, permettant ainsi de convertir les données pour les exposer sur le web.

### Le web de données par rapport aux bibliothèques :

La création du groupe d'incubation à l'intérieur du W3C, l'organisme de normalisation du web, s'est traduite par le besoin de faire connaître ces normes aux bibliothèques. Il s'agissait de créer une dynamique au niveau des bibliothèques par rapport aux normes et par rapport au web.

Le groupe de travail s'est réuni durant un an avec une réunion par semaine. Il s'agissait de réfléchir à la manière d'encourager l'utilisation des normes, avec l'interopérabilité au cœur des réflexions.

Le rapport final comporte parmi les livrables, des scénarios d'utilisation, la liste des données déjà disponibles et les vocabulaires. Ce vocabulaire englobe aussi celui des archives et des musées.

Le Linked open data n'a pas été étudié par le groupe, l'ouverture juridique n'étant pas l'objet de ses réflexions.

Le rapport répertorie les bénéfices attendus suivant les catégories de personnes :

- Pour les usagers : l'espace d'information n'est pas limité aux bibliothèques
- Pour les institutions : le décloisonnement de l'environnement de travail avec élargissement au niveau des outils, des processus.
- Pour les bibliothécaires : la rationalisation de la manière dont sont produites les données et la possibilité de se concentrer sur son domaine local.
- Pour les fournisseurs : des outils moins spécifiques dépassant le marché des bibliothèques.

Le rapport émet aussi des recommandations :

- Pour les décideurs : l'engagement des discussions sur les droits et l'ouverture des données. Il faut identifier ce qui peut être exposé sur le web très vite.
- Pour les normalisateurs : le développement de normes compatibles avec le Linked Data, l'intensification des échanges avec la communauté du web. Le but doit être l'utilisateur final.
- Pour les concepteurs : l'attribution d'URI, la préservation à long terme des vocabulaires RDF.
- Pour les bibliothécaires : le développement des politiques de préservation des vocabulaires.

Il faut prendre en compte que tous ces bénéfices et recommandations ont été émis au moment de la rédaction du rapport mais que la situation a évolué depuis.

Une traduction française est disponible à l'adresse suivante à l'initiative de la Cité de la musique :

<http://mediatheque.cite-musique.fr/MediaComposite/ARTICLES/W3C/XGR-Ild-fr.html>

D'autres actions sont menées autour du Linked Data :

- la communauté Dublin Core, très active
- la création d'un groupe au sein de l'IFLA sur le web sémantique.

### **Échanges autour du point 1. :**

G. Miura : pour se guider, les réalisations des bibliothèques pourront-elles servir d'exemples ou bien cela doit-il venir d'initiatives de réseaux ?

Les bibliothèques ont développé de petites choses, mais pas de méthodologie. Il s'agit encore de phase expérimentale. Mais le travail en réseau est indispensable pour avancer. Actuellement, il manque des propositions de nouveaux services à partir de l'utilisation de ces données.

Exemples de réalisations Linked data :

- la BBC,
- la BnF avec Data.bnf.fr,
- l'Abes avec IdRef
- Ouverture du Centre Pompidou virtuel, une plate-forme commune bibliothèque, musée, bases de données.

## **2 Présentation du rapport d'orientation demandé par le Comité stratégique bibliographique**

Ce rapport est demandé par les tutelles afin d'avoir une vision de tous les enjeux autour de RDA. Une version intermédiaire de ce rapport est présentée au Comité stratégique RDA. Le rapport final sera présenté au Comité stratégique bibliographique (CSB) qui se réunira à la BnF le 13 novembre.

Raymond Bérard, Gildas Illien, Françoise Leresche et Philippe Le Pape présentent l'avant projet du rapport.

## 2.1 Préambule, valoriser les données bibliographiques selon les usages

Ce préambule se veut un exposé de la problématique à laquelle les bibliothèques sont confrontées.

Au niveau des usages, on ne parle plus de catalogue mais de données. Les catalogues des bibliothèques sont sur le web mais pas encore dans le web. Aujourd'hui il n'est plus besoin de connaître les référentiels des bibliothèques pour trouver les documents.

Les données bibliographiques sont produites à des niveaux différents, ce qui nécessitera d'adapter en conséquence des services différents. De même, les données bibliographiques devront être liées aux documents numériques. Les données des bibliothèques devront pouvoir être partagées et enrichies.

Au niveau des collections, on parle de plus en plus de données et moins de collections. La structuration de l'information devient donc nécessaire.

Au niveau technologique, il s'agit d'adapter les standards du web et d'être indépendant des systèmes informatiques.

Au niveau de l'organisation, on se dirige vers la mutualisation et la coopération. La récupération doit s'effectuer de façon plus large et avec d'autres partenaires (les éditeurs, OCLC...). Il s'agit de rendre interopérables les différentes bases.

### Débat :

Y. Alix évoque la problématique de la différence d'organisation entre l'enseignement supérieur et la lecture publique, cette dernière se caractérisant par son hétérogénéité. La demande de données, liée à la récupération, est différente entre les deux secteurs. La demande de la lecture publique porte surtout sur des données minimales.

G. Illien informe du résultat de l'enquête menée par la BnF sur la récupération de notices. Elle montre d'une part que les réutilisateurs recherchent des notices de référence et de qualité, mais d'autre part que les moindres compétences en catalogage et la moindre familiarité avec les formats Marc des personnes récupérant les notices rendent malaisés les croisements et paramétrages nécessaires pour n'afficher que les informations qui leur sont utiles.

S. Mazens précise que du côté des BU la demande porte aussi sur des données minimales.

C. Jallet indique que le besoin de la lecture publique est la rapidité de mise à disposition de l'information. Le projet d'enrichissement se ferait via les lecteurs, les communautés du web. Le signalement de la production locale (comme les captations) se font hors catalogue.

Selon Y. Alix, il faudra définir où se trouve le niveau de granularité moyen.

S. Mazens précise que côté enseignement supérieur, le souci est justement de définir jusqu'où doit se faire l'accès : à l'article, au titre, à la base de données ? Actuellement la granularité est subie. La visibilité des catalogues sur le web est tout à fait pertinente, et la

structuration des données des bibliothèques est justement utile pour la connaissance de l'information.

Cet aspect devrait être intégré en préambule, ainsi que l'importance de la qualité des données et la disponibilité des données publiques.

Par ailleurs, la question de l'accès aux documents se pose. Les chercheurs n'ont pas toujours accès directement aux documents. Ils veulent accéder à un service (le PEB si l'accès au document en ligne n'est pas possible). La valeur du signalement s'en trouve encore plus importante.

G. Miura : La question de la temporalité n'est pas suffisamment explicitée. Elle demande à être précisée. Un calendrier plus global serait nécessaire pour organiser la décision.

R. Bérard cite le message du bureau de l'IABD : "il attire l'attention sur le fait que des spécificités qui ne concerneraient que quelques établissements ne doivent pas couper la France du reste du monde en renchérissant pour tous les établissements, jusqu'au plus petit, le coût des systèmes de gestion disponibles sur le marché français ou en rendant difficile ou plus lent leur adaptation à RDA".

G. Miura : Les questions à prendre en compte sont diverses : la FRBRisation, les licences et les outils. Les discussions avec les tutelles permettront d'établir le calendrier.

R. Bérard indique que la phase de transition va sans doute s'étendre sur une dizaine d'années. La temporalité est différente entre l'enseignement supérieur et les autres acteurs. Il peut être proposé une étude coté lecture publique sur le même schéma que celle menée coté enseignement supérieur. En tout cas, l'enseignement supérieur ne s'isolera pas par rapport aux acteurs internationaux.

R. Zacharyus signale l'attente des éditeurs par rapport à la réflexion engagée sur l'information bibliographique. Il pointe un risque de retard des éditeurs français de logiciels par rapport à leurs homologues étrangers.

Pour Françoise Leresche, la priorité est la mise en œuvre du modèle FRBR. L'utilisation du code de catalogage RDA sans FRBRisation des catalogues n'a pas d'utilité. Le format UNIMARC a évolué afin de prendre en compte le modèle FRBR.

Philippe Le Pape complète que les États-Unis ont un calendrier à 10 ans pour passer de MARC au web de données. Les deux problématiques sont différentes.

C. Jallet insiste sur l'accompagnement des collectivités et l'aide nécessaire des tutelles.

G. Illien : il faut repenser le rôle de la coopération bibliographique. Il y a matière à réinventer la coopération.

Y. Alix déclare que la chronologie se révèle plus importante que le calendrier. Un ordre logique doit être respecté.

G. Illien répond que les pré-requis sont importants (normatifs, techniques, juridiques). Par contre la chronologie ne sera pas identique selon les acteurs.

## **2.2 États des lieux : expérimentations en cours**

La BnF n'envisage pas de changer son système de production. Elle ne produira pas à court/moyen terme en RDA. Le système actuel a pensé les liens et la gestion des notices

d'autorité, les clés pour le futur. Il n'y a donc pas d'urgence à modifier le système actuel. La BnF s'est orientée vers la diffusion des données FRBRisées via *data.bnf.fr* et l'expérimentation *OpenCat*, qui vise à l'enrichissement des données locales par leur lien aux jeux de données culturelles de référence disponibles sur le Web. *OpenCat* est en cours de tests et sera présenté au salon du Livre 2013 et la documentation du projet sera publiée à l'été 2013.

La BnF regarde aussi les types d'utilisation des données (réutilisation commerciale, publique), dans l'objectif d'évaluer ces réutilisations pour proposer des services.

L'ABES indique que certaines de ses applications sont déjà dans le web de données (Calames, IdRef, Thèses.fr). Les données du Sudoc sont sorties en rdf mais non encore FRBRisées. L'ABES projette d'améliorer cette sortie.

L'ABES avait le projet avec OCLC et la DNB de la FRBRisation de la production du Sudoc. Mais les réponses à l'étude n'étaient pas suffisamment convaincantes, avec une incertitude sur les coûts engendrés. 2 solutions sont explorées en parallèle : à moyen terme la solution apportée par le futur système de gestion mutualisée ; à court terme FRBRiser indépendamment de tout outil comme cela a été fait pour les thèses de l'application *theses.fr*. Actuellement, l'ABES en est à l'étape d'analyse des données.

C. Jallet précise que la fourniture de données est vitale pour la lecture publique, pour le fonctionnement au quotidien et rappelle que le rôle de la BnF est primordial.

G. Illien indique que l'objectif est de mieux récupérer et envisage des entretiens avec les grands établissements.

## 2.3 Adopter RDA en France

F. Leresche présente les travaux engagés sur l'adoption de RDA en France. Le code de catalogage RDA a vocation à remplacer les AACR, règles de catalogage anglo-américaines. S'il se veut un code international, RDA demeure encore très anglo-saxon. Par ailleurs, les règles de RDA reflètent la situation des catalogues américains actuels et certaines d'entre elles illustrent une méconnaissance de la gestion des liens – même si l'ambition de RDA est de porter les catalogues sur le Web de données.

C'est pourquoi il a été décidé de ne pas adopter RDA en l'état car cela signifierait une régression par rapport à la situation actuelle des catalogues en France. Mais l'objectif demeure d'adopter RDA en France pour ne pas se couper de l'international.

Toutefois, la France a privilégié le niveau européen pour faire évoluer le code sur les points qui constituent des obstacles majeurs et s'est investie dans le groupe d'intérêt européen sur RDA (EURIG) où elle est un acteur décisif.

Les premières discussions sur les règles de RDA au niveau européen ont porté sur la mise en œuvre du modèle FRBR dans RDA, car cela a un impact fort sur la structuration des données. Par l'intermédiaire d'EURIG, le groupe technique a présenté au JSC des propositions d'évolution de RDA, et attend une réponse début 2013.

## 2.4 Les licences

La BnF et l'ABES s'accordent sur l'adoption de la licence ouverte de l'Etat mise au point par la mission Etalab, qui autorise tout type de réutilisation avec mention d'attribution. La décision reste à valider par le CA de l'ABES. La BnF utilise déjà cette licence ouverte pour ses données en RDF accessibles via *data.bnf.fr*. Un accord de principe a été pris pour

l'utilisation de la licence Creative Commons CC0, plus ouverte encore, pour les données en Dublin Core de Gallica exposées sur Europeana. Il reste à définir le régime des données en format MARC, pour lesquelles l'adoption de la licence ouverte permettrait une mise en cohérence des régimes juridiques de réutilisation des métadonnées. La question devrait être posée au prochain CA de la BnF.

Electre n'envisage pas pour l'instant de changement de sa politique sur les droits des notices.

Y. Alix rappelle que la notion de source est importante et qu'il serait incompréhensible qu'il n'y ait pas convergence entre les opérateurs

## **2.5 Enjeux du signalement de la documentation électronique**

Du côté de l'ABES, utilisation d'outils de nouvelle génération s'appuyant sur des bases de connaissance et non plus sur le catalogue traditionnel.

Du côté de la BnF, le catalogue ne peut pas évoluer pour la documentation électronique. Deux solutions sont possibles : s'inscrire dans le dispositif des Licences nationales, ou bien investir du côté du Dépôt légal.

Le dépôt légal de l'Internet n'est pas accessible en ligne pour des questions juridiques propres au dépôt légal, consultable à la BnF sur accréditation de recherche. Le signalement de certains jeux de données commence cependant à se faire, non pas par le catalogue de la BnF, mais directement sur *data.bnf.fr*. Suite au décret du 19/12/2011 encadrant le dépôt légal numérique, des discussions sont par ailleurs en cours avec le SNE pour le dépôt des fichiers numériques via l'extranet éditeurs.

S. Mazens demande quelle sera la temporalité du catalogage dans le SUDOC des livres et périodiques électroniques et quelle économie est attendue. Elle regrette l'absence de solutions intermédiaires, utilisables immédiatement par les BU.

R. Bérard rappelle la position du Conseil scientifique de l'ABES, qui n'est pas favorable à des solutions qui seraient chronophages sans garantie de résultat probant.

## **2.6 La formation**

Un plan de communication national est nécessaire, comprenant :

- une communication vers les décideurs
- une sensibilisation aux pré-requis et aux évolutions des pratiques en matière de catalogage, avec le soutien du CNFPT pour constituer un réseau. Mais la constitution de ce réseau est fragile et nécessite le soutien des directeurs, ainsi que des associations professionnelles.

Le Groupe technique a mis en place une formation de formateurs aux pré-requis à l'évolution des catalogues et des règles de catalogage (enjeux, web sémantique, FRBR, apport de RDA) : elle a réuni 26 formateurs, principalement des correspondants SUDOC, peu de formateurs venant de bibliothèques de lecture publique.

Des journées de sensibilisation sont prévues avec le CNFPT, à destination des directeurs d'établissement au 1er semestre 2013 (une journée dans (une journée dans chaque INSET - Institut spécialisé d'études territoriales). Ensuite des formations plus techniques sont envisagées.

Le Groupe technique prévoit aussi de développer la diffusion d'information sur les évolutions en cours par l'intermédiaire du site RDA-en-France. En particulier, une rubrique FAQ est en cours de préparation. Des relais d'information existent déjà à travers le blog de l'ABES ([rda@abes.fr](mailto:rda@abes.fr)) ou les pages professionnelles sur le site de la BnF, d'autres sont à mettre en place.

C. Jallet fait part de la différence d'information entre la lecture publique et les BU, et de son inquiétude sur le signalement du numérique hors catalogue.

Pour G. Illien, le lien entre catalogue et numérique existe. Ils peuvent se retrouver.

## **2.7 Évolution des SIGB**

La tendance qui se dessine, au moins en BU, est d'aller vers un système hébergé.

Du côté de l'enseignement supérieur, un groupe de travail sur un SIGB mutualisé a été mis en place, qui fédère BU et EPST. Fin 2013 sera présentée l'étude d'impact, l'appel d'offres pour un groupe d'établissements pilotes étant prévu à l'automne 2013.

La question est posée du rôle de la BnF par rapport aux bibliothèques de lecture publique qui ont besoin d'une tête de réseau.

M. Maisonneuve insiste sur le morcellement des bibliothèques de lecture publique, sans véritable relais fédérateur des associations. Les BDP peuvent jouer un rôle de mise en réseau.

La BnF insiste sur le rôle de la tutelle et sur l'aide qu'elle peut apporter. En prévision des évolutions vers le web de données, la BnF et l'ABES proposent de commander une étude d'impact sur les SIGB des bibliothèques de lecture publique, avec un regard plus ouvert, incluant des acteurs extérieurs au marché des SIGB.

La question de la sécurisation de l'espace public des données est évoquée : quelles données conserver dans le nuage public ? Quelles données l'ABES et la BnF peuvent elles mettre dans ce nuage ? Quel partage des rôles entre l'ABES et la BnF ?